A METHOD AND SYSTEM FOR EFFECTIVELY PERFORMING EVENT DETECTION IN A LARGE NUMBER OF CONCURRENT IMAGE SEQUENCES

Field of the Invention

The present invention relates to the field of video processing. More particularly, the invention relates to a method and system for obtaining meaningful knowledge, in real time, from a plurality of concurrent compressed image sequences, by effective processing of a large number of concurrent incoming image sequences and/or features derived from the acquired images.

Background of the Invention

Many efforts have been spent to improve the ability to extract meaningful data out of images captured by video and still cameras. Such abilities are being used in several applications, such as consumer, industrial, medical, and business applications. Many cameras are deployed in the streets, airports, schools, banks, offices, residencies – as standard security measures. These cameras are used either for allowing an operator to remotely view security events in real time, or for recording and analyzing a security event at some later time.

The introduction of new technologies is shifting the video surveillance industry into new directions that significantly enhance the functionality of such systems. Several processing algorithms are used both for real-time and offline applications. These algorithms are implemented on a range of platforms from pure software to pure hardware, depending on the application. However, these platforms are usually designed to simultaneously process a relatively small number of incoming image

sequences, due to the substantial computational resources required for image processing. In addition, most of the common image processing systems are designed to process only uncompressed image data, such as the system disclosed in U.S. Patent 6,188,381. Modern networked video environments require efficient processing capability of a large number of compressed video steams, collected from a plurality of image sources.

Increasing operational demands, as well as cost constrains created the need for automation of event detection. Such event detection solutions provide a higher detection level, save manpower, replace other types of sensors and lower false alarm rates.

Although conventional solutions are available for automatic intruder detection, license plate identification, facial recognition, traffic violations detection and other image based applications, they usually support few simultaneous video sources, using expensive hardware platforms that require field installation, which implies high installation, maintenance and upgrade costs.

Conventional surveillance systems employ digital video networking technology and automatic event detection. Digital video networking is implemented by the development of Digital Video Compression technology and the availability of IP based networks. Compression standards, such as MPEG-4 and similar formats allow transmitting high quality images with a relatively narrow bandwidth.

A major limiting factor when using digital video networking is bandwidth requirements. Because it is too expensive to transmit all the cameras all

the time, networks are designed to concurrently transmit data, only from few cameras. The transmission of data only from cameras that are capturing important events at any given moment is crucial for establishing an efficient and cost-effective digital video network.

Automatic video-based event detection technology becomes effective for this purpose. This technology consists of a series of algorithms that are able to analyze the camera image in real time and provide notification of a special event, if it occurs. Currently available event-detection solutions use conventional image processing methods, which require heavy processing resources. Furthermore, they allocate a fixed processing power (usually one processor) per each camera input. Therefore, such systems either provide poor performance due to resources limitation or are extremely expensive.

As a result, the needs of large-scale digital surveillance installations – namely, reliable detection, effective bandwidth usage, flexible event definition, large-scale design and cost, cannot be met by any of the current automatic event detection solutions.

Video Motion Detection (VMD) methods are disclosed, for example, in U.S. Patent 6,349,114, WO 02/37429, in U.S. Patent Application Publication 2002,041,626, in U.S. Patent Application Publication No. 2002,054,210, in WO 01/63937, in EP1107609, in EP1173020 , in U.S. Patent 6,384,862 , in U.S. Patent 6,188,381, in U.S. Patent 6,130,707, and in U.S. Patent 6,069,655. However, all the methods described above have not yet provided satisfactory solutions to the problem of effectively obtaining meaningful knowledge, in real time, from a plurality of concurrent image sequences.

It is an object of the present invention to provide a method and system for obtaining meaningful knowledge, from a plurality of concurrent image sequences, in real time.

It is another object of the present invention to provide a method and system for obtaining meaningful knowledge, from a plurality of concurrent image sequences, which are cost effective.

It is a further object of the present invention to provide a method and system for obtaining meaningful knowledge, from a plurality of concurrent image sequences, with reduced amount of bandwidth resources.

It is still another object of the present invention to provide a method and system for obtaining meaningful knowledge, from a plurality of concurrent image sequences, which is reliable, and having high sensitivity in noisy environments.

It is yet another object of the present invention to provide a method and system for obtaining meaningful knowledge, from a plurality of concurrent image sequences, with reduced installation and maintenance costs.

Other objects and advantages of the invention will become apparent as the description proceeds.

Summary of the Invention

While these specifications discuss primarily video cameras, a person skilled in the art will recognize that the invention extends to any appropriate image source, such as still cameras, computer generated images, pre-recorded video data, and the like, and that image sources

should be equivalently considered. Similarly, the terms video and video stream, should be construed broadly to include video sequences, still pictures, computer generated graphics, or any other sequence of images provided or converted to an electronic format that may be processed by a computer.

The present invention is directed to a method for performing event detection and object tracking in image streams. A set of image acquisition devices is installed in field, such that each device comprises a local programmable processor for converting the acquired image stream, that consists of one or more images, to a digital format, and a local encoder, for generating features from the image stream. The features are parameters that are related to attributes of objects in the image stream. Each device transmits a feature stream, whenever the number and type of features exceed a corresponding threshold. Each image acquisition device is connected to a data network through a corresponding data communication channel. An image processing server connected to the data network determines the threshold and processes the feature stream. Whenever the server receives features from a local encoder through its corresponding data communication channel and the data network, the server obtains indications regarding events in the image streams by processing the feature stream and transmitting the indications to an operator.

The local encoder may be a composite encoder, which is a local encoder that further comprises circuitry for compressing the image stream. The composite encoder may operate in a first mode, during which it generates and transmits the features to the server, and in a second mode, during which it transmits to the server, in addition to the features, at least a portion of the image stream in a desired compression level, according to

commands sent from the server. Preferably, each composite encoder is controlled by a command sent from the server, to operate in its first mode. As long as the server receives features from a composite encoder, that composite encoder is controlled by a command sent from the server, to operate in its second mode. The server obtains indications regarding events in the image streams by processing the feature stream, and transmitting the indications and/or their corresponding image streams to an operator.

Whenever desired one or more compressed image streams containing events are decoded by the operator station, and the decoded image streams are transmitted to the display of an operator, for viewing. Compressed image streams obtained while their local encoder operates in its second mode may be recorded.

Preferably, additional image processing resources, in the server, are dynamically allocated to data communication channels that receive image streams. Feature streams obtained while operating in the first mode may comprise only a portion of the image.

A graphical polygon that encompasses an object of interest, being within the frame of an image or an AOI (Area Of Interest) in the image may be generated by the server and displayed to an operator for viewing. In addition, the server may generate and display a graphical trace indicating the history of movement of an object of interest, being within the frame of an image or an AOI in the image.

The image stream may be selected from the group of images that comprises video streams, still images, computer generated images, and pre-recorded digital, analog video data, or video streams, compressed

using MPEG format. The encoder may use different resolution and frame rate during operation in each mode.

Preferably, the features may include motion features, color, portions of the image, edge data and frequency related information.

The server may perform, using a feature stream, received from the local encoder of at least one image acquisition device, one or more of the following operations and/or any combination thereof:

- License Plate Recognition (LPR);
- Facial Recognition (FR);
- detection of traffic rules violations;
- behavior recognition;
- fire detection;
- traffic flow detection;
- smoke detection.

The present invention is also directed to a system for performing event detection and object tracking in image streams, that comprises:

a) a set of image acquisition devices, installed in field, each of which includes:

a.1) a local programmable processor for converting the acquired image stream, to a digital format

a.2) a local encoder, for generating, from the image stream, features, being parameters related to attributes of objects in the image stream, and for transmitting a feature stream, whenever the motion features exceed a corresponding threshold;

b) a data network, to which each image acquisition device is connected through a corresponding data communication channel;

c); and

d) an image processing server connected to the data network, the server being capable of determining the threshold, of obtaining indications regarding events in the image streams by processing the feature stream, and of transmitting the indications to an operator.

The system may further comprise an operator display, for receiving and displaying one or more image streams that contain events, as well as a network video recorder for recording one or more image streams, obtained while their local encoder operates in its first mode.

Brief Description of the Drawings

The above and other characteristics and advantages of the invention will be better understood through the following illustrative and non-limitative detailed description of preferred embodiments thereof, with reference to the appended drawings, wherein:

Fig. 1 schematically illustrates the structure of a surveillance system that comprises a plurality of cameras connected to a data network, according to a preferred embodiment of the invention;

Fig. 2 illustrates the use of AOI's (Area of Interest) for designating areas where event detection will be performed and for reducing the usage of system resources, according to a preferred embodiment of the invention; and

Figs. 3A to 3C illustrate the generation of an object of interest and its motion trace, according to a preferred embodiment of the invention.

Detailed Description of Preferred Embodiments

A significant saving in system resources can be achieved by applying novel data reduction techniques, proposed by the present invention. In a

situation where thousands of cameras are connected to a single server, only a small number of the cameras actually acquire important events that should be analyzed. A large-scale system can function properly only if it has the capability of identifying the inputs that may contain useful information and perform further processing only on such inputs. Such a filtering mechanism requires minimal processing and bandwidth resources, so that it is possible to apply it concurrently on a large number of image streams. The present invention proposes such a filtering mechanism, called Massively Concurrent Image Processing (MCIP) technology that is based on the analysis of incoming image sequences and/or feature streams, derived from the acquired images, so as to fulfill the need for automatic image detection capabilities in a large-scale digital video network environment.

MCIP technology combines diverse technologies such as large scale data reduction, effective server design and optimized image processing algorithms, thereby offering a platform that is mainly directed to the security market and is not rivaled by conventional solutions, particularly with vast numbers of potential users. MCIP is a networked solution for event detection in distributed installations, which is designed for large scale digital video surveillance networks that concurrently support thousands of camera inputs, distributed in an arbitrarily large geographical area and with real time performance. MCIP employs a unique feature transmission method that consumes narrow bandwidth, while maintaining high sensitivity and probability of detection. MCIP is a server-based solution that is compatible with modern monitoring and digital video recording systems and carries out complex detection algorithms, reduces field maintenance and provides improved scalability, high availability, low cost per channel and backup utilities. The same

system provides concurrently multiple applications such as VMD, LPR and FR. In addition, different detection applications may be associated with the same camera.

MCIP is composed of a server platform with various applications, camera encoders (either internal or external to the camera), a Network Video Recorder (NVR) and an operator station. The server contains a computer that includes proprietary hardware and software components. MCIP is based on the distribution of image processing algorithms between low-level feature extraction, which is performed by the encoders which are located in field (i.e., in the vicinity of a camera), and high-level processing applications, which are performed by a remote central server that collects and analyzes these features.

The MCIP system described hereafter solves not only the bandwidth problem but also reduces the load from the server and uses a unique type of data stream (not a digital video stream), and performs an effective process for detecting events at real time, in a large scale video surveillance environment.

A major element in MCIP is data reduction, which is achieved by the distribution of the image processing algorithms. Since all the video sources, which require event detection, transmit concurrently, the required network bandwidth is reduced by generating a reduced bandwidth feature stream in the vicinity of each camera. In order to detect and track moving objects in digitally transmitted video sources by analyzing the transmitted reduced bandwidth feature, there is no need to

transmit full video streams, but only partial data, which contains information regarding moving objects.

By doing so, a significantly smaller data bandwidth is used, which reduces the demands for both the network bandwidth and the event detection processing power. Furthermore, if only the shape, size, direction of movement and velocity should be detected, there is no need to transmit data regarding their intensity or color, and thus, a further bandwidth reduction is achieved. Another bandwidth optimization may be achieved if the encoder in the transmitting side filters out all motions which are under a motion threshold, determined by the remote central server. Such threshold may be the AC level of a moving object, motion distance or any combination thereof, and may be determined and changed dynamically, according to the attributes of the acquired image, such as resolution, AOI, compression level, etc. Moving objects which are under the threshold are considered either as noise, or non-interesting motions.

One method for extracting features at the encoder side is by slightly modifying and degrading existing temporal-based video compressors which were originally designed to transmit digital video. The features may also be generated by a specific feature extraction algorithm (such as any motion vector generating algorithm) that is not related to the video compression algorithm. When working in this reduced bandwidth mode, the output streams of these encoders are definitely not a video stream, and therefore cannot not be used by any receiving party to produce video images.

Fig. 1 schematically illustrates the structure of a surveillance system that comprises a plurality of cameras connected to a data network, according to a preferred embodiment of the invention. The system 100 comprises n

image sources (in this example, n cameras, CAM1,....,CAMn), each of which connected to a digital encoder ENCj, for converting the images acquired by CAMj to a compressed digital format. Each digital encoder ENCj is connected to a digital data network 101 at point pj and being capable of transmitting data, which may be a reduced bandwidth feature stream or a full compressed video stream, through its corresponding channel Cj. The data network 101 collects the data transmitted from all channels and forwards them to the MCIP server 102, through data-bus 103. MCIP server 102 processes the data received from each channel and controls one or more cameras which transmit any combination of the reduced bandwidth feature stream and the full compressed video stream, which can be analyzed by MCIP server 102 in real time, or recorded by NVR 104 and analyzed by MCIP server 102 later. An operator station 105 is also connected to MCIP server 102, for real time monitoring of selected full compressed video streams. Operator station 105 can manually control the operation of MCIP server 102, whenever desired.

The MCIP (Massively Concurrent Image Processing) server is connected to the image sources (depicted as cameras in the drawing, but may also be any image source, such taped video, still cameras, video cameras, computer generated images or graphics, and the like.) through data-bus 103 and network 101, and receives features or images in a compressed format. In the broadest sense this is any type of network, wired or wireless. The images can be compressed using any type of compression. Practically, IP based networks are used, as well as compression schemes that use DCT, VideoLAN Client (VLC, which is a highly portable multimedia player for various audio and video formats as well as Digital Versatile Discs (DVDs), Video Compact Discs (VCDs), and various streaming protocols, disclosed in WO 01/63937) and motion estimation techniques such as MPEG.

The system 100 uses an optional load-balancing module that allows it to easily scale the number of inputs that can be processed and also creates the ability to remove a single point of failure, by creating backup MCIP servers. The system 100 also has a configuration component that is used for defining the type of processing that should be performed for each input and the destination of the processing results. The destination can be another computer, an email address, a monitoring application, or any other device that is able to receive textual and/or visual messages.

The system can optionally be connected to an external database to assist image processing. For example, a database of suspect, stolen cars, of license plate numbers can be used for identifying vehicles.

Fig. 2 illustrates the use of AOI's (Area of Interest) for reducing the usage of system resources, according to a preferred embodiment of the invention. An AOI is a polygon (in this Fig., an hexagon) that encloses the area where detection will occur. The rectangles indicate the estimated object size at various distances from the camera. In this example, the scene of interest comprises detection movement of a person in a field (shown in the first rectangle). It may be used in the filtering unit to decide if further processing is required. In this case, the filtering unit examines the feature data. The feature stream is analyzed to determine if enough significant features lie within the AOI. If the number of features that are located inside the AOI and comprise changes, exceeds the threshold, then this frame is designated as possibly containing an event and is transferred for further processing. Otherwise, the frame is dropped and no further processing is performed.

The MCIP server receives the reduced bandwidth feature stream (such a feature stream is not a video stream at all, and hence, no viewable image can be reconstructed thereof) from all the video sources which require event detection. When an event is detected within a reduced bandwidth stream that is transmitted from a specific video source, the central server may instruct this video source to change its operation mode to a **video stream mode**, in which that video source may operate as a regular video encoder and transmits a standard video stream, which may be decoded by the server or by any receiving party for observation, recording, further processing or any other purpose. Optionally the video encoder also continues transmitting the feature stream at the same time.

Working according to this scheme, most of the video sources remain in the **reduced bandwidth mode**, while transmitting a narrow bandwidth data stream, yet sufficient to detect events with high resolution and frame rate at the MCIP server. Only a very small portion of the sources (in which event is detected) are controlled to work concurrently in the **video stream mode**. This results in a total network bandwidth, which is significantly lower than the network bandwidth required for concurrently transmitting from all the video sources.

For example, if a conventional video surveillance installation that uses 1000 cameras, a bandwidth of about 500Kbp/s is needed by each camera, in order to transmit at an adequate quality. In the reduced bandwidth mode, only about 5Kbp/s is required by each camera for the transmission of information regarding moving objects at the same resolution and frame rate. Therefore, all the cameras working in this mode are using a total bandwidth of 5Kbp/s times 1000 = 5Mbp/s. Assuming that at steady state

suspected objects appear in 1% of the cameras (10 cameras) and they are working in video stream mode, extra bandwidth of 10 times 500Kbp/s = 5Mbp/s is required. Thus, the total required network bandwidth using the solution proposed by the present invention is 10Mbp/s. A total required network bandwidth of 500Mbp/s would be consumed by conventional systems, if all the 1000 cameras would concurrently transmit video streams.

The proposed solution may be applicable not only for high-level moving objects detection and tracking in live cameras but also in recorded video. Huge amounts of video footage are recorded by many surveillance systems. In order to detect interesting events in this recorded video, massive processing capabilities are needed. By converting recorded video, either digital or analog, to a **reduced bandwidth stream** according to the techniques described above, event detection becomes much easier, with lower processing requirements and faster operation.

The system proposed in the present invention comprises the following components:

1. One or more MCIP servers
2. One or more dual mode video encoders, which may be operated at **reduced bandwidth mode** or at **video stream mode, according to remote instructions.**
3. Digital network, LAN or WAN, IP or other, which establishes communication between the system components.
4. One or more operator stations, by which operators may define events criteria and other system parameters and manage events in real time.

5. An optional Network Video Recorder (NVR), which is able to record and play, on demand, any selected video source which is available on the network.

Implementation for security applications:

Following is a partial list of types of image processing applications which can be implemented very effectively using the method proposed by the present invention:

Video Motion Detection – for both indoor and outdoor applications. Such application is commonly used to detect intruders to protected zones. It is desired to ignore nuisances such as moving trees, dust and animals. In this embodiment of the present invention manipulates input images at the stream level in order to filter out certain images and image changes. Examples of such filtering are motion below a predetermined threshold, size or speed related filtering all preferably applied within the AOIs, thus reducing significantly the amount of required system resources for further processing. Since the system is server-based and there is no need for installation of equipment in the field (except the camera), this solution is very attractive for low budget application such as in the residential market.

Exceptional static objects detection -. this application is used to detect static objects where such objects may require an alarm, By way of example, such objects may comprise an unattended bag at the airport, a stopped car on a highway, a person stopped at a protected location and the like. In this embodiment the present invention manipulates the input images at the stream level and examines the motion vectors at the AOIs. Objects that stopped moving are further processed.

License Plate Recognition. - this application is used for vehicles access control, stolen or suspected car detection and parking automation. In this embodiment, it is possible to detect wanted cars using hundreds or more cameras installed in the field, thus providing a practical detection solution.

Facial Recognition - this application is desired for biometric verification or detection device, for tasks such as locating criminals or terrorists and for personal access control purposes. Using this embodiment offers facial recognition capability to many cameras in the field. This is a very useful tool for large installations such as airports and public surveillance.

Smoke and flames detection - this application is used for fire detection. Using this embodiment of the invention, all the sites equipped with cameras may receive this service in addition to other application without any installation of smoke or flame detectors.

Traffic violations - this application detect a variety of traffic violation such as red light crossing, separation line crossing, parking or stopping at forbidden zone and the like. Using this embodiment, this functionality may be applied for many cameras located along roads and intersections, thus significantly optimizing police work.

Traffic flow analysis - this application is useful for traffic centers by automatically detecting any irregular traffic events such as traffic obstacles, accidents, too slow or too fast or too crowded traffic and the like. Using this embodiment, traffic centers may use many cameras located as desired at the covered area in order to provide a significantly better control level.

Suspicious vehicle or person tracking - this application is used to track objects of interest. This is needed to link a burglar to an escape car, locate a running suspect and more. Using this embodiment, this functionality may be associated with any selected camera or cameras in the field.

It should be noted that each of those applications or their combination may each be considered as a separate embodiment of the invention, all while using the basic structure contemplated herein, while specific embodiments may utilize specialized components. Selection of such component and the combination of features and applications provided herein is a matter of technical choice that will be clear to those skilled in the art.

Figs. 3A to 3C illustrate the generation of an object of interest and its motion trace, according to a preferred embodiment of the invention. Fig. 3A is an image of a selected AOI (in this example, an elongated zone, in which the presence of any person is forbidden), on which the MCIP server 102 generates an object, which is determined according to predefined size and motion parameters, received from the corresponding encoder. The object encompasses the body of a person, penetrating into the forbidden zone and walking from right to left. The motion parameters are continuously updated, such that the center of the object is tracked. The MCIP server 102 generates a trace (solid line) that provides a graphical indication regarding his motion within the forbidden zone. Fig. 3B is an image of the same selected AOI, on which the MCIP server 102 generates the object and the trace (solid line) that provides a graphical indication regarding his motion within the forbidden zone from left to right and more closely to the camera. Fig. 3C is an image of the same selected AOI, on which the MCIP server 102 generates the object and the trace (solid line) that provides a graphical indication regarding his motion within the

forbidden zone again from right to left and more closely to the camera. The filtration performed by the corresponding encoder prevents the generation of background movements, such as tree-tops and lower vegetation, which are considered as background noise.

The above examples and description have of course been provided only for the purpose of illustration, and are not intended to limit the invention in any way. As will be appreciated by the skilled person, the invention can be carried out in a great variety of ways, employing more than one technique from those described above, all without exceeding the scope of the invention.